

GOVERNMENT AI IN PRACTICE

Research and analysis from the ThinkCapital GIAG Initiative

ISSUE 3 · LATE APRIL 2026



EARLY SIGNAL: FROM THE RESEARCH

Agencies presenting complete AI governance documentation and agencies with governance programs that have genuine operational influence look identical in a compliance review. The GIAG research is designed to measure the difference.

The implementation fidelity question is not whether the RMF has been adopted. It is whether adoption has changed anything.

This distinction between framework as compliance artifact and framework as operational governance discipline is the central question the GIAG research is designed to investigate comparatively across agency types.

From the Editor

Issue 2 examined the compliance deadline built into OMB M-25-21 and what it does and does not measure. The response from practitioners confirmed something the research was already pointing toward: the agencies most at risk are not the ones failing to produce documentation. They are the ones producing documentation confidently, without a clear view of whether it reflects how they authentically govern AI.

Two developments since Issue 2 sharpen that concern. First, the pace of AI capability deployment continued to outrun policy. The models that federal agencies are now being asked to govern are meaningfully more capable than the models that OMB guidance was written to address. Second, a series of high-profile disputes between federal agencies and AI vendors surfaced exactly the governance failure this research was designed to examine: when the relationship deteriorated, neither side had a written governance architecture capable of resolving the disagreement. Verbal assurances and informal understandings are not governance. That distinction became very expensive.

Conversations with practitioners working in administrative governance roles across organizational types add a parallel that belongs in this issue. Framework adoption is broad. The evidence that adoption has changed decision behavior is rare. The structural condition is the same regardless of sector: governance requirements exist, compliance activity runs on event-driven cycles rather than continuous operations, and no standard measurement practice distinguishes organizations that are governed from organizations that are documented. Government is where the stakes are highest.

This issue explains what Stream One is designed to find, introduces a new analytical frame on governance timing that practitioners have found useful, and makes the clearest case yet for why this research needs direct participation from the people running these programs.

The Implementation Variance NIST Doesn't Measure

The NIST AI Risk Management Framework has achieved something few voluntary government frameworks manage: genuine adoption. Federal Chief AI Officers reference it. Agency IT governance policies cite it. OMB M-25-21 effectively anchors federal AI governance requirements to it. State CIOs have adopted it as the de facto reference architecture even where no state mandate exists.

What the framework does not do, and what it was not designed to do, is distinguish between agencies that have implemented it as a governance discipline and agencies that have implemented it as a compliance artifact.

The compliance documentation exists. The governance record is harder to find.

The RMF's four core functions, Govern, Map, Measure, and Manage, are defined in terms of activities and documentation. An agency can satisfy every documentation requirement across all four functions while operating a governance program that has never changed a deployment decision, never stopped a system from going live, and never produced a risk assessment that led to a concrete action. The framework is silent on that gap. Compliance with the framework as written does not require demonstrating that the framework has done anything.

This is not a criticism of NIST. The RMF is a reference architecture, not an audit standard. It was never intended to be self-enforcing. The practical consequence is that a federal agency with complete, well-documented AI RMF implementation and an agency whose governance program has genuine operational influence look identical in a compliance review. They do not perform identically under pressure.

Two kinds of records get produced by federal AI governance programs. The first documents that a governance process ran: artifacts were produced, reviews were completed, authorization was granted. The record is accurate and auditable. The second documents that a governance process worked: risk characterizations were updated as the system evolved, scope changes triggered reviews, oversight mechanisms caught something and changed a decision that would otherwise have gone through. The first kind of record is what almost every current federal AI audit mechanism is designed to verify. The second kind is what governance is supposed to produce. No standard methodology exists for distinguishing between them.

Agencies generating the first kind of record report the same compliance score as agencies generating the second. That measurement gap is the central research question for GIAG Stream One.

The Four Implementation Patterns

Based on the GIAG Stream One research design and preliminary practitioner conversations, four distinct implementation patterns characterize how agencies deploy the RMF in practice. These are descriptive categories, not evaluative rankings.

Pattern	Label	Operational characteristic
A	Documentation Compliance	RMF present as policy and documentation discipline. Governance activity clusters around IG and audit cycles. AI deployments proceed on their own operational momentum between reviews.

Pattern	Label	Operational characteristic
B	Selective Adoption	Substantive engagement with Govern and Map; Measure and Manage nominally present. Risk identification occurs. Risk response is inconsistent. Framework functions operate independently rather than in sequence.
C	Adapted Implementation	Deliberate, documented adaptations to the RMF based on mission context. Framework treated as a starting point, not a checklist. Governance activity is not primarily audit-driven.
D	Measurement-Grounded Governance	Measurement infrastructure built into RMF implementation. Risk assessments connected to deployment decisions. The framework has demonstrably changed outcomes: systems modified, constrained, or declined based on governance findings.

Pattern A produces full compliance documentation. So does Pattern D. The operational difference between them is invisible to documentation review alone.

Three Governance Clocks, All Running Behind

Government AI governance has a timing problem, and it is not one problem but three. Three distinct clocks govern AI deployment in federal agencies. None of them run at the speed of the systems they are governing. None of them are synchronized with each other.

The Assessment Clock

AI systems receive a governance assessment before deployment. After authorization, assessment cadences follow audit cycles, budget cycles, or annual review schedules. These cycles were designed for systems that change slowly. AI deployments, particularly agentic systems in live workflows, change quickly. By the time the next scheduled assessment occurs, the system being assessed is often not the system currently running. The documentation describes a prior state.

The Policy Clock

OMB guidance, agency AI use policies, and RMF implementation plans move on bureaucratic timelines. The policy governing an active deployment was often written before the deployment existed, or before the capability it addresses was fully understood. M-25-21 and M-25-22 are currently driving compliance activity across federal agencies. The agentic deployment patterns those memos were designed to govern are already more advanced than the policy architecture that describes them.

The Accountability Clock

When something goes wrong, the accountability mechanism activates: incident review, inspector general inquiry, GAO audit, congressional attention. This clock runs slowest of all, often by years. It operates on the documentation that existed at authorization, not at the time of the incident. The gap between what was authorized and what was operating when the failure occurred is rarely the first question asked. It should be the first question asked.

All three clocks are asynchronous with each other and with AI operational tempo. The governance structures agencies are building were designed for a slower clock than the systems they are governing.

This is not an argument for less governance. It is an argument for governance that runs on the system's clock, not the agency's administrative calendar. The agencies operating at Pattern C and Pattern D in the implementation typology have figured out some version of this. The research is designed to document exactly what they figured out.

When Governance Architecture Becomes a Contract Dispute

Two high-profile federal agency and AI vendor conflicts over the past several months produced the same instructive result. In both cases, the dispute was framed publicly as a guardrails disagreement. In both cases, the actual problem was that governance architecture had never been made explicit. When the relationship deteriorated, both sides discovered that what had looked like shared understanding was a set of assumptions that neither party had written down.

The most instructive detail from the second dispute: a competing vendor reached agreement with the same agency on terms that included the same two restrictions the original vendor had been holding firm on. The substantive governance positions were identical. The outcomes were not. That means the dispute was never about the guardrails. It was about who had the authority to insist on them, and whether that authority had been established clearly enough in advance to survive a deteriorating relationship.

Three questions every government AI deployment contract should answer before it is signed:

- Where does autonomous operation end and human accountability begin, stated explicitly and agreed by both parties?
- What triggers a scope review when a use case expands beyond its original parameters?
- What is the dispute resolution mechanism when governance terms are contested, and who has the authority to invoke it?

Organizations that answer those questions before deployment will be in a fundamentally different position than those who answer them during a crisis.

The governance question is not which vendor to choose. It is whether the terms governing the partnership, including what happens when those terms are disputed, are **explicit enough to survive a change in political conditions**. Right now, in most federal AI deployments, they are not.

What the GIAG Stream One Research Is Seeking

Stream One conducts structured 30-minute interviews with government IT leaders who have direct experience with AI deployment, risk management, or governance program design. Six dimensions of implementation are examined in each conversation.

Framework scope

Which RMF functions are applied, and with what fidelity to the full framework? What adaptations have been made, and were they deliberate or ad hoc?

Measurement infrastructure

What metrics, if any, assess governance quality between audit cycles? How does the agency know its RMF implementation is working? Can the agency produce any evidence that governance activity changed a deployment outcome?

Decision influence

Has the RMF governance process ever produced a deployment decision: a delay, a modification, a constraint, or a halt? If so, what drove it? If not, what would need to be present for it to do so? This is the single most diagnostic question in the research protocol.

Governance triggers

What initiates a governance review? Audit deadlines and IG visits are common answers. The research is also looking for agencies where operational signals, including model updates, use case expansion, and incident reports, trigger reviews independently of the compliance calendar.

Cross-functional accountability

Who owns the AI governance function: the CIO, the Chief AI Officer, the CISO, or some combination? Where does accountability sit when a governance question crosses those boundaries? In many agencies, the answer is: whoever is closest to the problem at the time.

Agentic AI governance

For agencies with agentic AI deployments or active evaluation programs: how does the RMF framework apply, and where does it break down? The framework was designed primarily for defined-scope AI systems. Agentic deployments introduce governance requirements the current framework does not fully address. Stream Two examines that question directly.

Conversations are conducted under a standard disclosure and confidentiality protocol. Participants choose their level of attribution, from named contributor to fully anonymized, at the point of consent. All participants receive a summary of research findings before publication. This is a professional peer conversation, not a compliance review.

Early Signal

The GIAG research streams are now actively recruiting participants. Three directional signals have emerged from initial practitioner conversations and from the extended public record on NIST AI RMF implementation to date. These are observations, not findings. They become findings when tested against a structured practitioner sample.

The audit-cycle dependency is nearly universal

Across agency types and sizes, governance activity concentrates around IG reviews and compliance deadlines. Continuous governance operating on its own operational signals is the exception. This pattern is not surprising for a framework in its early adoption phase. What it suggests is that the distinguishing variable for Pattern C and Pattern D agencies is not framework sophistication but organizational governance culture: someone decided that governance should run on the system's clock, not the audit schedule.

Government agencies are not uniquely susceptible to this condition. Conversations with practitioners in administrative and operational governance roles across sectors surface the same structural gap: framework requirements exist, compliance activity is event-driven rather than continuous, and no standard measurement practice distinguishes organizations that are actively governed from organizations that are fully documented. The check-boxing problem is not a government IT phenomenon. It is a governance design phenomenon that appears consistently wherever frameworks are adopted without measurement infrastructure to verify that the framework is doing anything. Government is arguably where the stakes are highest.

Measurement infrastructure is the rarest capability

The presence of pre-deployment metrics and post-deployment monitoring that feeds back into governance decisions is the characteristic most consistently absent in current implementations. Agencies know what they deployed. They are less consistently able to demonstrate what the deployment produced

in terms of risk outcomes. The absence of that feedback loop is what makes it structurally impossible to tell the difference between a governance program that is working and one that is producing records.

The Chief AI Officer role is not yet anchored

The M-25-21 requirement to designate Chief AI Officers has been broadly met. What the research is beginning to surface is that the CAIO role is institutionally young and variably positioned. In some agencies it carries genuine cross-functional authority. In others it is a designation without a governance function. Where the CAIO sits relative to the CIO and CISO, and whether those three functions coordinate or operate in parallel, appears to be a significant predictor of governance quality. The agencies where the CAIO has formal authority to delay or modify a deployment decision are not the same agencies where the CAIO produces an annual AI inventory report.

A governance title is not the same as governance function. The research is measuring the difference.

Five Questions the Field Is Not Asking

The volume of AI governance commentary has increased sharply in 2026. Most of it addresses what agencies should be doing. Very little of it addresses whether what agencies are currently doing is producing the outcomes governance is supposed to produce. The GIAG research is built around the second question. These are the five diagnostic questions that current policy and audit practice are not requiring agencies to answer.

Has your governance process ever stopped anything?

Not delayed it pending documentation. Not flagged it for monitoring. Stopped it, modified it in a material way, or declined to authorize it. If the answer is no, the governance program may be producing records rather than decisions. That is a testable condition. It is not currently being tested.

What would your governance program look like six months after the last audit?

Governance activity that concentrates around audit preparation and dissipates afterward is not operational governance. It is audit preparation. The question of what continuous governance looks like between compliance deadlines is not addressed in any current federal AI audit framework.

Who holds accountability when a scoped use case expands?

Agentic AI systems in operational workflows do not stay bounded. Use cases expand, integrations deepen, and the system that was authorized is regularly not the system currently running. The governance question is not whether expansion happens. It is who is responsible for detecting it and what authority they have to act on it. Most agency governance programs have not answered this question.

Can you distinguish governance from documentation after the fact?

If an incident occurred today with an AI system operating in your agency, could you produce evidence showing that the governance program had been monitoring the condition that produced the incident? Or would you produce evidence showing that a governance process ran and documentation was filed? Those are different answers to the same question. The audit mechanism does not currently require agencies to know the difference.

What does your governance program do that a checklist does not?

A checklist confirms that required activities occurred. A governance program produces decisions. If the output of your governance program is primarily documentation of activity, the program is functioning as a checklist. That is worth knowing. It is especially worth knowing before an incident makes it visible.

The Participation Ask

The Stream One research needs practitioners who have built, evaluated, or advised on AI governance programs in government agencies. The most valuable perspectives are not the ones that reflect a successful Pattern D implementation. The most valuable perspectives are the ones that reflect the full range: agencies working through early-stage compliance, agencies that have hit governance gaps in practice, and agencies that have had to make hard calls on AI deployment based on governance findings.

If you are a federal or state agency IT leader, a government AI program manager, a Chief AI Officer, or a practitioner who advises agencies on AI governance or NIST RMF implementation, your experience is directly relevant to what this research is designed to answer. If you read the five questions above and found that one of them does not have a clean answer at your agency, that is precisely the conversation this research needs.

The conversation takes 30 minutes and is conducted under your choice of attribution terms. You do not need to represent a governance success story.

To express interest or ask a question before committing: thinkcapital.org/research.html

What's Ahead

Issue 4 (June 2026) will report on the GIAG research intake: the participant profile taking shape, the implementation patterns emerging across the early interview set, and the first cross-agency observations on where RMF implementation variance is most pronounced. That issue is contingent on interview completion. The research is on track to begin structured interviews in May.

Working Paper 1, Implementation Fidelity, is in active development. The paper presents the full Stream One findings and the implementation typology with supporting evidence. Target release is Q3 2026.

Stream Two, Human Oversight Quality, remains active. Issue 5 will cover that stream's research design and participation requirements. Practitioners with experience governing agentic AI deployments in government contexts are the target audience for that stream. If that description fits you, the conversation does not have to wait for Issue 5.

Government AI in Practice

Published by ThinkCapital LLC · Belmont, California · Michael Bragen, Principal
thinkcapital.org · thinkcapital.substack.com · michael.bragen@thinkcapital.org

© 2026 ThinkCapital LLC. All rights reserved.